

## **Theory and Practice of Data Assimilation in Ocean Modeling**

Robert N. Miller

College of Oceanic and Atmospheric Sciences

Oregon State University

COAS Admin. Bldg. 104

Corvallis, OR 97331-5503

Phone: (541) 737-4555 Fax: (541) 737-2064 Email: [miller@coas.oregonstate.edu](mailto:miller@coas.oregonstate.edu)

Award Number: N00014-04-1-0262

### **LONG-TERM GOALS**

The long-range goal of this project is to form the best picture of the ocean as an evolving system based on data assimilation, i.e., the construction of a composite estimate of the state of the ocean based on a combination of observed data with computational model output, and to use that picture to understand the physical processes that govern the ocean's behavior. Oceanic observations are sparse and models are limited in accuracy, but taken together, one can form a quantitative description of the state of the ocean that is superior to any based on either models or data alone. Along with the goals of analysis and prediction, we seek reliable estimates of the errors in our results. We expect our results to have implications beyond the technical challenges of data assimilation. In particular, we believe this research will lead to enhanced understanding of the implications of nonlinearity and randomness for predictability of the ocean and atmosphere.

In keeping with our goal of providing reliable error estimates for our data assimilation products, we seek to develop efficient methods for estimating useful statistical measures of errors in stochastic forecast models, and information about stochastic systems is contained in the associated probability density function (PDF). The PDFs of nonlinear stochastic models are not, in general, Gaussian, so we must find methods for forecast evaluation based on information about the particular PDF generated by the model.

Since our goal is the development of practical analysis and forecast systems for the ocean, we want to solve remaining scientific problems involved in transition from data assimilation experiments tuned to specific models and data sets to operational analysis and prediction on a research basis. This will involve rigorous quantification of the information content of each data set, as well as quality control, a problem with which the ocean modeling community has limited experience.

### **OBJECTIVES**

The principal objective of this project is the development, implementation and evaluation of practical data assimilation methods for regional to basin scale ocean models. Since data assimilation methods that give the most and best information are highly resource intensive, and often not practical for use

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>2009</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2009 to 00-00-2009</b>	
4. TITLE AND SUBTITLE <b>Theory and Practice of Data Assimilation in Ocean Modeling</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Oregon State University, College of Oceanic and Atmospheric Sciences, COAS Admin. Bldg. 104, Corvallis, OR, 97331-5503</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>7</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

with detailed models, we are particularly interested in the price paid in terms of accuracy and confidence for using economical but suboptimal data assimilation methods.

Direct calculation of full PDFs is not feasible for practical models of the ocean or atmosphere, but useful approximations to the PDF can be calculated from Monte-Carlo experiments, by virtue of the fact that the number of truly independent degrees of freedom in practical models is very much smaller than the dimension of the state vector. This intuition is the motivation for the ensemble methods that have become popular in recent years. Our experience with Monte-Carlo methods in simplified systems has led us to investigate the details of methods for ensemble generation that have been presented in the community. The motivation for these specialized methods for generating ensembles is precisely the specification of the PDF of a complex model whose behavior is believed to be captured by a relatively small number of independent degrees of freedom. By detailed study of the behavior of ensembles in increasingly complex models, we hope to gain the insights necessary to generate the most efficient ensembles, which should, in turn, lead to the error estimates necessary for data assimilation systems and prior estimates of forecast accuracy.

Optimized methods require accurate knowledge of the statistics of the errors in the model and the data. It is therefore an objective to understand in detail the sensitivity of the data assimilation scheme to the details of the defining error estimates.

## **APPROACH**

The basic assumptions underlying data assimilation methods in use or proposed are known to be false to some degree. We plan to study the consequences of these assumptions by constructing a hierarchy of schemes with decreasing reliance on ad hoc assumptions. It is our guiding philosophy that the best way to learn how to design and implement the most economical methods that meet our needs is to begin by implementing methods which are as close to optimal as possible. From that point, we can quantify the loss of accuracy and the saving of resources associated with each simplification of the model or the data assimilation scheme.

Work is proceeding toward a theoretical basis for the next generation of data assimilation methods in which randomness and nonlinearity must be taken into account. To this end, we are applying tools from stochastic differential equations and from dynamical systems theory. Since our model systems are characterized by high dimensional state spaces, Monte Carlo methods must be used to study the behavior of the stochastic systems.

The theory of nonlinear filtering provides a framework in which problems of data assimilation with nonlinear models and non-Gaussian noise sources can be treated (see, e.g., Miller et al., 1999). In the case of linear models and Gaussian noise sources, this theory reduces to the familiar Kalman filter. In the formal theory of nonlinear filtering, the final result is not a single model state vector or trajectory in state space, but a PDF defined as a scalar function of the state variables and time. From this PDF, the mean, median, mode, or other statistic can be computed for use as the working estimate of the state of the system, along with the desired confidence intervals. The assignment of confidence limits corresponds in the case of a group of particles in physical space to drawing contours in the spatial domain which can be expected to define a region which contains, say, 90% of the particles.

The problem is that for even schematic models of the ocean or atmosphere, an unrealistically large number of particle trajectories in phase space must be calculated in order to represent the PDF

faithfully. Useful ensemble analysis therefore requires judicious choice of ensemble members. We have concentrated our recent efforts on evaluation of ensemble methods, which we see as facilitating the generation of the forecast error estimates necessary for data assimilation. These forecast error estimates are of interest in and of themselves, since they have the potential of providing a priori estimates of the reliability of a given forecast.

Results from the theory of dynamical systems lead to methods for explicit construction of the low dimensional spaces in which meaningful probabilistic calculations can be performed on complex systems. We are now finishing our work on a local model of the Kuroshio, and have begun to extend it to a model of the Pacific basin. The simplest of our models is a regional two-layer quasigeostrophic model that reproduces the observed bimodality. It operates on a state space with several thousand dimensions. This is two orders of magnitude greater than that of earlier schematic models, and, for this reason alone, presents significant technical challenges.

We now have a basis of comparison with more complex models, up to and including eddy resolving primitive equation models of the north Pacific. We are now in the process of applying our methods from dynamical systems and stochastic calculus to a suite of models, in order to understand propagation of errors and the evolution of the PDF arising from random initial and boundary conditions in a state space of workable dimension. This should allow us to construct reliable data assimilation systems for use with simulated and real data from the Kuroshio. In a parallel effort, we are using multivariate statistical techniques to isolate relevant low-dimensional subspaces of the state spaces of detailed models.

Many different models, based on fundamentally different physical assumptions, exhibit the observed bimodality of the Kuroshio in some form. We are now in the process of comparing our model to different models and to observed data in order to determine a basis for distinction among the physical mechanisms in the different models.

Technical support for this project is provided by Ms. Laura Ehret.

## **WORK COMPLETED**

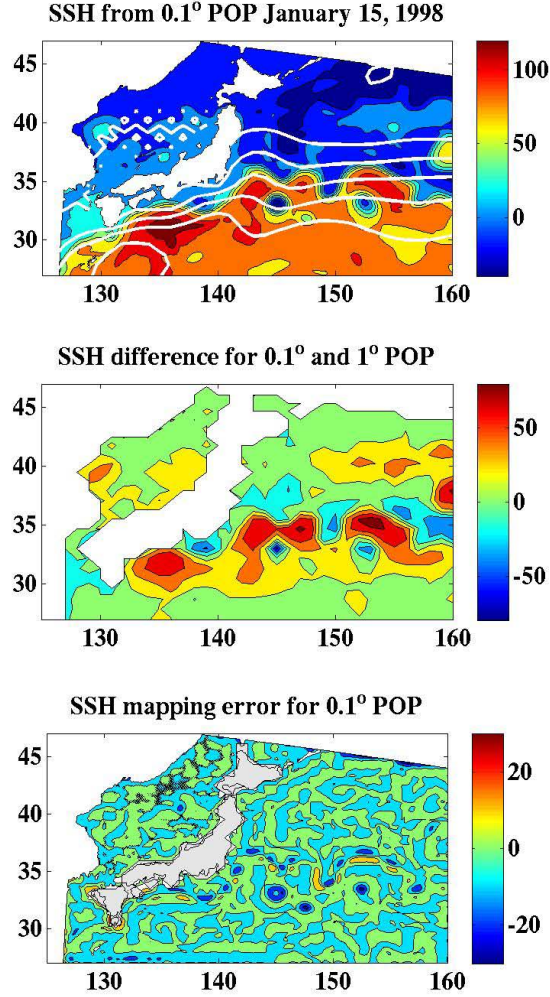
We have categorized the representation error in a coarsely resolved model of the north Pacific in comparison with an eddy resolving model, and we have formulated and verified the basis of an ensemble generation method that takes the physical limitations of the model into account. We have evaluated our representation error calculations by generating simulated fields of SST representation error according to our statistics.

## **RESULTS**

As expected, in formulating a data assimilation scheme for a non-eddy-resolving model, much of the variability in the model data misfit must be assigned to representation error. Previous authors (e.g., Cane et al., 1996, Desroziers et al. 2001 and Janic and Cohn, 2006) have characterized representation error as a consequence of interpolation error. In practice, the difficulty encountered by coarsely resolved models in reproduction of the details of intense currents and other characteristic ocean features lies in the physical approximations that they must employ. This is illustrated in figure 1, which depicts, through comparison of results from an  $0.1^\circ$  model (Smith et al., 2000) and a  $1^\circ$  model, an example of the difference between the consequences of interpolation error and physical error. The

height difference across the Kuroshio is similar for both models, but the Kuroshio in the  $1^\circ$  model has a width of almost  $8^\circ$  compared to the narrow 100 km of the output of the  $0.1^\circ$  model. The middle panel shows the SSH difference between the two model results. The scales of the anomalies associated with meanders and eddies are resolved on the  $1^\circ$  grid, but the model physics in the  $1^\circ$  model do not generate the instabilities responsible for the characteristic scales of Kuroshio eddy and meander variability. From the bottom panel, which shows the interpolation error obtained by averaging the  $0.1^\circ$  model on the  $1^\circ$  grid and remapping back to the  $0.1^\circ$  grid, we see that the interpolation error is much smaller in amplitude ( $\sim 20$  cm) and horizontal scale ( $< 1^\circ$ ) than the SSH differences between the  $0.1^\circ$  and  $1^\circ$  models.

Leading EOFs of the computed representation error have their greatest weights in places such as the Kuroshio where the model cannot be expected to reproduce observations faithfully. Results of data assimilation experiments with a multivariate optimal interpolation method based on our error estimates show relatively little impact of assimilation of SST and SSH. This is because the model is reasonably accurate at simulating those phenomena for which its dynamics resemble those found in nature. Details are presented in Richman and Miller (2009).



**Figure 1.** Difference between an eddy-resolving 0.1° model and the coarse resolution 1° model in this study in the vicinity of the Kuroshio. Top panel: SSH for January 15, 1998, with contours of the coarse resolution model SSH overlain. Middle panel: SSH difference between the two models. Bottom panel: Interpolation error obtained by averaging the 0.1° model on the 1° grid and remapping back to the 0.1° grid. From Richman and Miller, 2009.

## IMPACT/APPLICATIONS

Major weather centers, including the US National Center for Environmental Prediction (NCEP) and the European Center for Medium-Range Weather Forecasting (ECMWF) now use ensemble methods to evaluate the reliability of operational forecasts; see Molteni et al. (1996), Toth and Kalnay (1993). Our work on Monte-Carlo methods should provide enhanced capability for evaluation of forecasts of the ocean and atmosphere, in addition to application to data assimilation. Our work on breeding modes and planned work on other schemes for ensemble generation should provide significant guidance in optimizing methods for ensemble generation.

Our work on estimation of representation error statistics and statistics of model error that take physical model limitations into account should lead to new efficient ensemble generation methods in two ways. Ensembles of model forecasts informed by the ability of the model to represent physical variability can be constructed, as can ensembles of simulated representation error fields generated by stochastic models of representation error (cf. Richman and Miller, 2009). Ensembles of model forecasts, combined with ensembles of simulated representation error can be combined to provide fields of simulated data suitable for OSSEs or for interdisciplinary modeling and data assimilation.

## TRANSITIONS

We are working with scientists at NCEP to begin the process of incorporating our error estimates into their operational climate forecast system.

## RELATED PROJECTS

Estimating the representation error of satellite and in-situ data for data assimilation into ocean models.

Particle Filters and Ecological Models (PFEM): Application of chainless Monte-Carlo methods to mapping the ecology of the North Pacific Ocean

## REFERENCES

Cane, M. A., A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert and A. J. Busalacchi, 1996: Mapping tropical Pacific sea level: Data assimilation via a reduced state Kalman filter. *J. Geophys. Res.*, **101**, 22,599–22,617.

Desroziers, G., O. Brachemi and B. Hadamache, 2001: Estimation of representativeness error caused by the incremental formalism of variational data assimilation. *Q. J. R. Meteorol. Soc.*, **127**, 1775–1794.

Janic, T. and S. E. Cohn, 2006: Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Mon. Wea. Rev.*, **134**, 2900–2915.

Miller, R. N., E. F. Carter and S. T. Blue, 1999: Data assimilation into nonlinear stochastic models. *Tellus*, **51A**, 167–194.

Molteni, F., R. Buizza, T. N. Palmer and T. Petroliaxis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Q. J. R. Meteorol. Soc.*, **122**, 73–119.

Richman, J. G. and R. N. Miller, 2009: Model representation error for ocean data assimilation. *Ocean Modelling*, submitted.

Smith, R. D., Maltrud, M. E., Bryan, F. O. and Hecht, M. W., 2000: Numerical simulation of the north Atlantic at  $1/10^\circ$ . *J. Phys. Oceanogr.* **30**, 1532–1560.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.*, **74**, 2317–2330.

## **PUBLICATIONS**

Richman, J. G. and R. N. Miller, 2009: Model representation error for ocean data assimilation. *Ocean Modelling*.